



Formato para registro de Unidades de aprendizaje 2021

I.- Datos de identificación de la unidad de aprendizaje

Unidad académica:	Multisede (CIC, CIDETEC, ESCOM, ESFM, UPIITA)									
Programa académico:	Maestría en Ciencia y Tecnología de Inteligencia Artificial y Ciencia de Datos									
		Doctorado					Orientación profesional			
	x	Maestría				x	Orientado a la investigación			
		Especialidad					Con la industria			
							Especialidad médica			
Nombre de unidad de aprendizaje:	Sesión de colegio donde se propuso:						Fecha de propuesta:			
	Procesamiento de Lenguaje Natural									
Tipo de unidad de aprendizaje:	Clave de la unidad de aprendizaje:						Créditos:		5 <i>REP 2017</i>	
	Semanas del semestre		18		Horas a la semana:		4		Horas totales: 72	
	Obligatoria:				Optativa:		x		Observaciones:	
	Semestre:		1-3							
	Teórica (%):				Práctica (%):				Teórico-prácticas (%): 100	
Área del conocimiento:	Ingeniería y Ciencias Fisicomatemáticas		x		Ciencias Sociales y Administrativas				Ciencias Médico Biológicas	
									Interdisciplinario	
Modalidad no escolarizada:	No escolarizada				Nombre de la Plataforma:					
	Mixta				Presencial (%):				En plataforma (%):	
Horas establecidas en el programa de estudios:	Presenciales (si procede) (horas x semana)						En plataforma (horas x semana):			



Formato para registro de Unidades de aprendizaje 2021

I. Aprendizajes que el estudiante deberá demostrar al finalizar

Conocimientos	Habilidades y destrezas	Actitudes y valores
- El alumno será capaz de realizar análisis del lenguaje a diversos niveles: léxico y sintáctico - El manejo de la ambigüedad a distintos niveles - La habilidad de clasificar textos mediante algoritmos simples	- Identificar claramente las ventajas y desventajas de cada uno de los conocimientos adquiridos. - Capacidad para extrapolar los conocimientos adquiridos.	- Deseo por investigar por su cuenta de los temas vistos en clase. - Capacidad para comprender y reflexionar sobre los temas del curso.

Resolución que aborda la propuesta con su enfoque disciplinar

La meta de la Lingüística Computacional es que las computadoras puedan desarrollar tareas útiles que involucren el lenguaje humano, tareas como habilitar la comunicación humano-máquina, mejorar la comunicación humano-humano o hacer simplemente procesamiento de texto y del habla. En la era de la información en línea, los individuos e instituciones cada vez se enfrentan más a grandes cantidades de texto que son críticas para el desempeño de las tareas relevantes; al mismo tiempo, fenómenos que antes no eran detectables pueden ser estudiados gracias al análisis automático de esas grandes cantidades de texto. Para ello, en este curso el alumno aprenderá herramientas estadísticas necesarias, que comprenden los enfoques cuantitativos al procesamiento automático del lenguaje.

II. Proximidad formativa

Áreas multi, inter y transdisciplinarias	Líneas de Generación y Aplicación de Conocimiento	Sectores sociales
- Ciencias Cognitivas Computacionales - Procesamiento de Lenguaje natural	- Aprendizaje automático - Redes neuronales y aprendizaje profundo - Reconocimiento de patrones - Minería de datos, descubrimiento de conocimiento y analítica avanzada - Minería de texto y procesamiento de lenguaje natural	Mejora de habilidades en la implementación de soluciones, que tomen en cuenta a los diversos sectores de la población, considerando las implicaciones que éstas pueden llegar a tener.
Estrategia de asociación:		



Formato para registro de Unidades de aprendizaje 2021

Se buscará aplicar las técnicas aprendidas en la solución de problemas que enfrentan organismos gubernamentales y no gubernamentales cuya actividad principal está asociadas a resolver problemas asociados a temas de interés.

III Metodología de enseñanza – aprendizaje

Descripción
Las unidades se complementan con material audiovisual para que los estudiantes adquieran los conceptos presentados en cada unidad. Durante el desarrollo del curso se implementarán las técnicas vistas por cada unidad. Dichas implementaciones serán consideradas para la evaluación, con un valor de 60% Asimismo, al final de cada unidad, se realizará un examen de conocimientos, con valor de 20% Asistencia y exposiciones de los alumnos contarán como el 20% restante.

Evidencias como proceso de aprendizaje	Evidencias integradoras (resultados que contribuyen al curriculum)	Ponderación
Algoritmos implementados, resúmenes realizados y mapas mentales de los temas. Presentación de proyectos relacionados con los temas vistos en clase y los temas de interés del estudiante.		

IV. Descripción de la participación esperada en el estudiante

Receptiva	Resolutiva	Autónoma	Estratégica
-----------	------------	----------	-------------



Formato para registro de Unidades de aprendizaje 2021

--	--	--	--

Contenido temático

1. Introducción al Procesamiento de Lenguaje Natural (PLN) (2H)

- 1.1 Definición
- 1.2 Aplicaciones
- 1.3 Herramientas y recursos

2. Preparación de datos textuales (4H)

- 2.1 Adquisición de Datos
- 2.2 Normalización de Texto
 - 2.2.1 Tokenización
 - 2.2.2 Lematización y Stemming
 - 2.2.3 Eliminación de elementos no relevantes
 - 2.2.4 Etiquetado de categorías gramaticales

3. Representación Vectorial (6H)

- 3.1 Bolsa de Palabras
- 3.2 Codificación de Vector Uno-a-Uno
- 3.3 TF-IDF
- 3.4 Incrustaciones de Palabras
 - 3.4.1 Word2Vec



Formato para registro de Unidades de aprendizaje 2021

3.4.2 Glove

3.4.3 ElMo

4. Extracción de aspectos semánticos (10H)

4.1 Modelado del Lenguaje con n-gramas

4.1.1 Evaluación de Modelos de Lenguaje

4.1.2 Suavizado

4.2 Asociación entre palabras

4.3 Similitud de texto

4.4 Palabras claves

4.5 Modelado de tópicos

5. Aprendizaje profundo para el procesamiento de lenguaje natural (18H)

5.1 Redes neuronales recurrentes (RNN)

5.1.1 Aplicaciones de las RNN en el PLN

5.1.2 Arquitecturas de las RNN

5.1.3 Redes LSTM y GRU

5.1.4 Modelo codificador-decodificador

5.2 Modelo transformador

5.2.1 Mecanismos de atención

5.2.2 Preentrenamiento

5.2.3 Codificador bidireccional

5.2.4 Ajuste fino

5.2.5 Aprendizaje por transferencia

6. Aplicaciones de PLN (30H)

6.1 Corrección de texto

6.1.1 Ortográfico

6.1.2 Gramatical

6.1.3 Estilo

6.2 Clasificación de textos



Formato para registro de Unidades de aprendizaje 2021

6.3 Análisis de sentimientos 6.4 Respuesta a preguntas 6.5 Chatbots y sistemas de diálogo 6.6 Traducción automática 6.7 Resumen automático 7. Consideraciones éticas para PLN (2H)

V. Secuencia programática

N o .		T e m a		Objetivo de aprendizaje / competencia específica		Tiempo/Horas/Semanas		
						8		
Actividad(es):		No. Descripción de las actividades: Contenidos:				Tipo de interacción(es):		
						Referencias (s):		
Evidencia(s):								

VI. Habilitadores tecnológicos

Disposiciones		Especificaciones / descripción de efectos
x	Conectividad	
x	Habilidades digitales	



Formato para registro de Unidades de aprendizaje 2021

x	Interoperabilidad	
x	Datos abiertos	
x	Servicios de nube	
x	Aprendizaje automático	
	Simulación	
	Realidad aumentada	
	Otro...	

VII. Referencias

Conferencias magistrales

1.	
2.	
3.	

Notas complementarias

Documentales / electrónicas

1. Kumar, Ela. (2011) Natural language processing. IK International Pvt Ltd.
2. Calvo, Hiram. (2013) Procesamiento práctico de lenguaje natural, Editorial SMIA.
3. Gelbukh, A., & Calvo, H. (2018) Automatic syntactic analysis based on selectional preferences (Vol. 765). Springer.
4. Haro, S. N. G., & Gelbukh, A. (2007) Investigaciones en análisis sintáctico para el español. Instituto Politécnico Nacional, Dirección de Publicaciones.
5. Kamath, Uday (2019) Deep learning for NLP and speech recognition. Springer.
6. Mathangi Sri. (2021) Practical natural language processing with python. Apress.
7. L. Ashok Kumar, D. Karthika Renuka (2023) Deep learning approach for natural language processing, speech, and computer vision. CRC Press



Formato para registro de Unidades de aprendizaje 2021

VIII. Créditos y responsabilidades

Responsabilidad	Nombre completo	Clave de nombramiento /No. de empleado
Coordinador (Autor)	Francisco Hiram Calvo Castro	9016-EC-12
Participante (Coautor)	Idalia Maldonado Castillo	
Participante (Coautor)	Joel Omar Juárez Gambino	
Participante (Coautor)	Hind Taud	13748-EE-18
Participante (Coautor)		
Asesor didáctico / Diseñador Instruccional		
Tecnólogo educativo / Comunicólogo		
Corrector de estilo		
Programador multimedia / Diseñador gráfico		
Otro...		

VERIFICACIÓN GENERAL DE LA PLANEACIÓN DIDÁCTICA

Por la División de Operación y Promoción al Posgrado de la SIP

Nombre _____

FIRMA _____

REVISIÓN DE LA PLANEACIÓN DIDÁCTICA (VIABILIDAD)

Por la Subdirección de Diseño y Desarrollo de la DEV

Nombre _____

FIRMA _____



Formato para registro de Unidades de aprendizaje 2021

VERIFICACIÓN PARA SU PUESTA EN OPERACIÓN	REVISIÓN TÉCNICO-PEDAGÓGICA PARA LA MODALIDAD
Por la Dirección de Posgrado	Por la Dirección para la Educación Virtual
Nombre _____	Nombre _____
FIRMA _____	FIRMA _____
SELLO DE VALIDACIÓN	